DS-GA 1013: Project
Intelligible Principal Component Analysis

Kho, Lee
ltk224

Peeters, Jonas
jp5642

April 4, 2020

Deliverables

1. Load the data
2. Performance benchmarking
3. PCA
4. Thresholding PCA
5. Simple Principal Components
6. CPEV versus Sparsity plots
7. SCoTLASS
8. SPCA

Simple Principal Components

Notes

Vines (2000) generalized this approach without restricting the integer set solution using a sequence of simplicity preserving transformations. Each transformation chooses a pair of vectors and rotates them orthogonally in such a way that the variance associated with the currently higher variance component of the pair is increased at the expense of the lower variance component

This method what introduced by Vines (2000). We transform the set of orthogonal axes. Essentially we will be orthogonally rotating and rescaling two orthogonal vectors d_1 and d_2. In general we can write an orthogonal rotation as $f_1 = Pd_1$ where

$$P = \begin{bmatrix} 1 & \|d_2\|_2^2\beta \\ \beta & -\|d_1\|_2^2 \end{bmatrix} \tag{1}$$

The new variance-covariance matrix V^* of the rotated data can be found as $V^* = P^T V P$ where V is the variance-covariance matrix of the data before rotation.

$$V = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} \quad \text{and} \quad V^* = P^T V P = \begin{bmatrix} 1 & \|d_2\|_2^2\beta \\ \beta & -\|d_1\|_2^2 \end{bmatrix}^T \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} \begin{bmatrix} 1 & \|d_2\|_2^2\beta \\ \beta & -\|d_1\|_2^2 \end{bmatrix}$$
$$= \begin{bmatrix} h \end{bmatrix}$$

The variance in the direction f_1 can be found as follows:

$$\text{Var}(Xf_1) = f_1^T V f_1 = (Pd_1)^T V P d_1 = d_1^T V^* d_1 \tag{2}$$

References

Vines, S. K. (2000, jan). Simple principal components. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(4), 441–451. doi: 10.1111/1467-9876.00204